A CULTURE-FREE PERFORMANCE TEST OF GENERAL LEARNING ABILITY

Peter Adams Young

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

A CULTURE-FREE PERFORMANCE TEST
OF GENERAL LEARNING ABILITY

by

Peter Adams Young

Thesis Advisor: J. K. Arima

December 1975

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Culture-Free Performance Test of General Learning Ability		5. TYPE OF REPORT & PERIOD COVERED Masters Thesis 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Peter Adams Young		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		12. REPORT DATE December 1975 13. NUMBER OF PAGES 77
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Naval Postgraduate School Monterey, California 93940		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Discrimination Learning
Information Theory
Human Learning
Rote Learning
Culture-Free Test
Learning Aptitude
Performance Testing
Racial Differences

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Random polygons were used as stimuli in a two-choice multiple discrimination learning paradigm designed to test individual learning ability. Information processing rate (IPR) was used as the measure of learning ability. Variables in the test design were: racial group (white, nonwhite), pacing mode (self-paced, machine-paced), and stimulus similarity. Subjects were 121 white and 39 nonwhite male Navy

recruits. Over 10 trials, a learning effect was demonstrated, with internal (split-half) test reliability of .84 overall. White performance was superior to nonwhite only in the machine-paced mode. Significant correlation between IPR and Navy General Classification Test (GCT) scores was seen for the entire group, but was present only in the white subgroup when the sample was divided by race. Stimulus similarity did not prove to be a significant factor. It was concluded that a reliable, culture-free test of general learning ability was practicable, although its validity with respect to on-job performance has yet to be established.

A Culture-Free Performance Test

of General Learning Ability

by

Peter Adams Young

Lieutenant, United States Navy

B.S., United States Naval Academy, 1967

Submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL

ABSTRACT

Random polygons were used as stimuli in a two-choice multiple discrimination learning paradigm designed to test individual learning ability. Information processing rate (IPR) was used as the measure of learning ability. Variables in the test design were: racial group (white, nonwhite), pacing mode (self-paced, machine-paced), and stimulus similarity. Subjects were 121 white and 39 nonwhite male Navy recruits. Over 10 trials, a learning effect was demonstrated, with internal (split-half) test reliability of .84 overall. White performance was superior to nonwhite only in the machine-paced mode. Significant correlation between IPR and Navy General Classification Test (GCT) scores was seen for the entire group, but was present only in the white subgroup when the sample was divided by race. Stimulus similarity did not prove to be a significant factor. It was concluded that a reliable, culture-free test of general learning ability was practicable, although its validity with respect to on-job performance has yet to be established.

TABLE OF CONTENTS

LIST OF TABLES

TABLE

LIST OF FIGURES

FIG.

I. STATEMENT OF THE PROBLEM

Personnel selection and classification have long been areas of vital interest to the military. The accelerating advance of weapons technology in this century has vastly increased the personnel requirements of all branches of the armed forces. Such advance, when coupled with recent reductions in force and manpower levels, have served to place yet more emphasis on selection and training.

Entry of the United States into World War I, with its concomitant massive personnel classification and training requirements, saw the development of and introduction of the first group intelligence tests designed for military use--the Army Alpha and Beta Tests. Nearly two million men were given these tests in the course of the war, and these results provided much of the data base for studies of ethnic, racial, and other cultural differences in intelligence and ability in subsequent years (Matarazzo, 1972). New demands for skilled manpower brought about by the outbreak of World War II led to the adoption of the Army General Classification Test (AGCT), also a group intelligence test designed specifically for the military.

Postwar developments and refinements in the field of military personnel testing and classification followed the general form of the earlier efforts. Emphasis on group testing of general intelligence and abilities continued, as

evidenced by the heavy use of the Armed Forces Qualification Test (AFQT) by all services until only very recently. Current policy is typified by the Navy's Basic Test Battery (BTB), which seeks to measure not only general intelligence, but also arithmetic reasoning ability and aptitude in specific areas such as mechanics and clerical work.

Assignment of Navy recruits to technical school training is currently made primarily on the basis of performance on certain component parts of the BTB, and failure to attain the requisite "cutoff" scores for a given school is basis for denial of advanced training in that specialty. The Navy maintains an ongoing study of the validity of BTB scores as predictors of school performance (Thomas, 1972a, 1972b).

Recent emphasis on racial and cultural imbalances in group tests of intelligence and abilities, prompted in part by Federal legislation designed to eliminate irrelevant bias, has led to renewed investigation of all aspects of personnel testing and selection. In addition, efforts to upgrade the overall quality of Navy personnel in the face of force and manpower cutbacks and the loss of the draft have pointed up a new approach to the problem with emphasis shifting to human development and training rather than selection alone. It can be anticipated that the current objective of a smaller, better-trained Navy in the near future will only increase the demand for adequate prediction of performance in training and on the job. Inherent in this demand is the minimization of needless losses to the selection program of people who may be

capable and trainable, but who lack the verbal or cultural background necessary to good performance on group-administered, paper-and-pencil, intelligence or aptitude tests.

A great deal of controversy and theory surrounds the discussion of the nature of human intelligence and innate ability. Efforts by Binet early in this century to quantify intellectual development levels led to the definition of the Intelligence Quotient (IQ) by Wilhelm Stern in 1914. In subsequent years, a variety of tests of human intelligence has emerged. While on the whole valid predictors of academic performance, most of these tests rely on an individual's level of intellectual development as a basis for determining "intelligence" (Matarazzo, 1972). Variance in environmental or cultural opportunity within the United States renders such measurement of intelligence vulnerable to the criticism of racial or cultural bias. Placement of emphasis on acquired knowledge in measuring intellect will invariably result in continued questioning of the validity of those measurements when applied to disadvantaged segments of the population.

Resolution of this problem is complicated by the interaction between inherent ability and environmental opportunity in determining an individual's intellectual development. While inherited or innate ability sets limits on this development, exposure to environmental factors which foster growth determines to a large degree the level actually attained. Cattell's (1963) definition of fluid and

crystallized intelligence, Hebb's (1972) treatment of intelligence A and intelligence B, and Jensen's (1968) definition of Level I and Level II intelligence exemplify recent attempts to explore the original ideas of inherited and acquired intelligence.

The impact of cultural differences on general intelligence tests is easily perceived, if not measured. "Achievement" testing of intelligence is seen as susceptible in many ways (and varying degrees) to these differences, and the resultant bias in scoring can lead to over- or under-prediction of performance or aptitude for minority groups (Thomas, 1972c). The existence of cultural or racial bias in the Navy's BTB has been identified by "in house" study (Stephan, 1973; Thomas, 1972c). Nonetheless, the BTB is maintained as the primary enlisted personnel classification tool, largely due to demonstrated high validity in predicting technical school grades (Thomas, 1972a, 1972b). Current school assignment policies within the Navy show, however, increasing concern with utilization of qualified minority personnel. This concern is echoed by recent emphasis on minority recruiting throughout the nation. This avowed objective of attracting and training qualified minority group people places still heavier demands on the selection and classification processes in the Navy to be both valid and unbiased.

The key to successful utilization of personnel within the highly technical environment of today's military lies in

training. The worth of an individual to the Navy can be directly tied to his or her ability to acquire the knowledge and skills of a given job specialty or rating. More generally, this can be interpreted as the ability to learn. The sole purpose behind development and administration of the BTB is in predicting performance in a training environment in the hope that this performance will relate to actual job performance in the Fleet. While a great deal of data has been collected on school performance and initial screening scores on selection tests (Thomas, 1972a, 1972b), measurement of actual job performance in the Navy has proved extremely difficult. In addition, those job performance measurements which have been made to date have relied upon supervisory rating of performance, a methodology which has shown very little validity in other studies by other services (Fox, et al., 1969).

Basic Test Battery scores are indicative however, not of aptitude for training so much as acquired knowledge or experience. In addition, emphasis is on verbal or academic material such as found in the school environment. The impact of this emphasis is not readily apparent until viewed in light of the heavy reliance on on-the-job training (OJT) in the Navy.

Completion of advanced technical schooling is, of course, a prerequisite for successful performance of complex technical tasks in the Navy, but almost without exception, extensive OJT is necessary before an individual can perform his or her task effectively in the Fleet. The major portion of this training

is conducted under actual operational conditions, with emphasis on learning-while-doing and observation of skilled technicians at work. Accordingly, a selection device must accurately predict ability to learn in this environment as well as in the school situation.

In summary, then, it would appear that critical issues in the Navy's selection program center about three key areas:

(1) Response to the characteristics and composition of the current and projected recruit manpower pools,

(2) Development of selection devices or concepts which reflect the native ability of the individual, and

(3) Emphasis on learning and performance in an operational environment rather than in the "schoolhouse" alone.

It would seem that some general indicator of an individual's ability to learn would be of benefit to the Navy's search for valid predictors of job performance. Indeed, a recent recruiting document lists "good ability to learn" or "above average learning ability" as requirements for success in Navy technical specialties (USN, 1973). Where success is dependent upon training, the ability to respond to this training--in nonverbal as well as verbal areas--stands as an essential attribute to be measured.

The primary objective of this study was to design, construct, and evaluate an objective test of individual learning

ability using nonverbal instruments. The guiding concepts for the test dictated that it be relatively easy to administer, as culture-free as possible, and as far removed as possible from current paper-and-pencil "aptitude" intelligence tests now in use. The resultant test was conceptualized as a nonverbal supplement to the current test battery, not as a replacement for established tests for academic aptitude.

II. TEST DEVELOPMENT

A. DISCRIMINATION LEARNING AS THE TEST TASK

Discrimination Learning (DL) tasks have for many years been used as fundamental tests of intellectual development levels. While rooted in animal behavior study, DL has been used in countless studies of human learning processes. Extensive use of DL in the fields of developmental and abnormal psychology has provided a basis for its application in studies of adult human learning processes as well.

The relative simplicity of the majority of existing DL tests and techniques (owing largely to the design of such tests for animals, children, or retardates) renders them inapplicable to the measuring of adult human learning ability (Green and O'Connell, 1969). Nonetheless, the basic nature of a DL test--that of a performance test that relies upon the ability to learn to distinguish one item from another--justifies investigation into its possible applications to testing human intellect.

Tests of job-related skills, while necessarily performance tests in themselves, as a rule provide only a narrow view of a testee's aptitude. Little or no attempt is made to measure actual intelligence or learning ability. Rather, these tests tend to be oriented toward measures of physical or perceptual motor skills peculiar to a certain task or field. While the narrow field of concentration of such tests

improves their validity within that field, few are generally applicable to a wide range of skills or training programs.

Introduction of DL test techniques into this area of personnel selection would provide measures of learning ability heretofore unavailable in reliable form. While tests in use today can provide accurate indication of an individual's tactile sensitivity, kinesthetic sense, dexterity, reaction speed, etc., no absolute measure of learning ability is found that is based on actual performance testing.

Recent investigation of verbal DL by Grey (1971), Baltutis (1972), Arima and Grey (1972a and 1972b), Bugarin (1973), and Arima (1974) provided a great deal of insight into the dynamics of serial DL of sets of verbal stimuli by adults.

These studies employed the premises of information theory as it relates to information presentation rate and information content. The use of verbal material as stimuli in these studies severely restricts their applicability to a test of general learning ability. The highly cultural and cognitive aspects of verbal materials render them virtually unusable in cases where subjects are drawn from a culturally diverse population.

Further, strong scientific evidence of physiological specialization within the human brain suggests that verbal material is not processed in the same fashion nor in the same brain areas as nonverbal material (Ornstein, 1972). Indeed, even the memory process may display this same specialization, reserving one brain center for the retention of visual

(pictorial, scenic) information, and another for processing linguistic or verbal material (Haber, 1970). Tests involving verbal stimuli and/or processing, including many paper-and-pencil tests currently in use, simply do not reach a great proportion of an individual's abilities. Others, such as pattern analysis tests, may only touch on these areas. Yet many of these abilities are vital to the effective performance of tasks in the Navy.

A return to nonverbal stimuli and responses, as employed in a great amount of DL experimentation, is thus seen as imperative if the ideal of a truly culture-free test of native learning ability is to be maintained. The problem is then perceived as that of developing a nonverbal replica of the models used in earlier studies of verbal discrimination learning. In this framework, DL tasks become more complex in that multiple discriminations must be learned concurrently, making the total task more applicable to the measurement of human learning ability.

B. TEST CONSTRUCTION

Construction of a nonverbal DL test suitable for administration to a culturally diverse population of human adults begins with the investigation of the information content involved. Verbal DL studies demonstrated the importance of the information presentation rate in the learning process (Baltutis, 1972; Bugarin, 1973).

1. Information Measurement Considerations

Quantification of the information presentation rate is possible in the case of DL in that stimuli are presented in discrete categories. Application of information theory to a set of discrete choices between discrete stimuli provides an absolute measure of the amount of information contained in each choice. When the initial probability of choice for each alternative in a DL stimulus set is known, information theory permits measurement of the reduction of response uncertainty over repeated exposures to a given stimulus set. In the case where each alternative in a stimulus set is equally likely to be selected, the information content of that set is given by:

$$I = \log_2 (N)$$

where

I = information content in bits

N = number of alternatives (choices) in a stimulus set

Thus a stimulus set containing two equally likely choices (items) contains one bit of information, a four-choice set two bits, and so forth. Stating the concept in other words, a subject can be said to process one bit of information when he selects an alternative from a set of two equally likely stimuli presented simultaneously. The mental process implied in this activity is the reduction of uncertainty involved in choosing the correct stimulus for response.

2. The Verbal DL Model

Gray (1971) directly related learning speed with the rate of presentation of stimulus information, and the subsequent work by Baltutis (1972) and Bugarin (1973) further

confirmed the relationship between IPR and learning performance in a verbal DL task situation. Of greater importance to this study, however, is the implication that the verbal DL model can be applied to the measurement of general learning ability.

Construction of a visual DL test along the lines of a verbal DL test necessarily centered about the location and selection of suitable stimuli. The nature of the basic test model required a fairly large number of distinguishable stimuli that were as free of cultural influence or implications as possible.

3. Stimulus Materials

The basic discrimination requirement for the test was determined to be that of shape or pattern discrimination. Avoidance of physiological complications, such as color blindness, further restricted the nature of the stimuli by eliminating size and color as discrimination factors. For these reasons, two-dimensional, black-and-white patterns of uniform size were investigated.

The need for a relatively long list of distinguishable shapes would eliminate the use of basic geometric shapes as used in many other visual DL experiments. The desire to avoid culturally-oriented stimuli would also eliminate employment of so-called "familiar objects."

Fitts, et al., (1956) investigated the implications and construction of metric histoforms. This work was paralleled by that of Attneave and Arnoult (1956) in that both

teams were concerned with the generation and informational aspects of random two-dimensional figures. It was felt that this area provided the greatest promise of suitable stimuli. Preliminary research into these forms has led to the selection of a set of 30, two-dimensional, metric polygons generated by the method of Attneave and Arnoult (1956) and listed in a study by Arnoult (1956). These items are presented in Figure 1.

Evaluation of the information content, subjective similarity, and other attributes of the polygons was deemed necessary prior to actual construction of stimulus lists to be used in testing. Of major concern were the possible effects of intra- and inter-item similarity among figures, resemblance to familiar objects (association value), relative complexity, etc., as well as possible unforeseen preferences on the part of any subject for a given polygon over another in a forced-choice situation. This concern was generated by the dependence of the information content of a choice situation on the probability of selection of one item over another.

In order to gain some insight into as many of these factors as possible, an initial experiment was conducted. The purpose of this preliminary study was to discover any significant tendency on the part of a group of subjects to choose one stimulus item over another upon initial naive exposure to a pair of polygons. In addition, some measure of the degree of subjective similarity between items presented in pairs was sought.

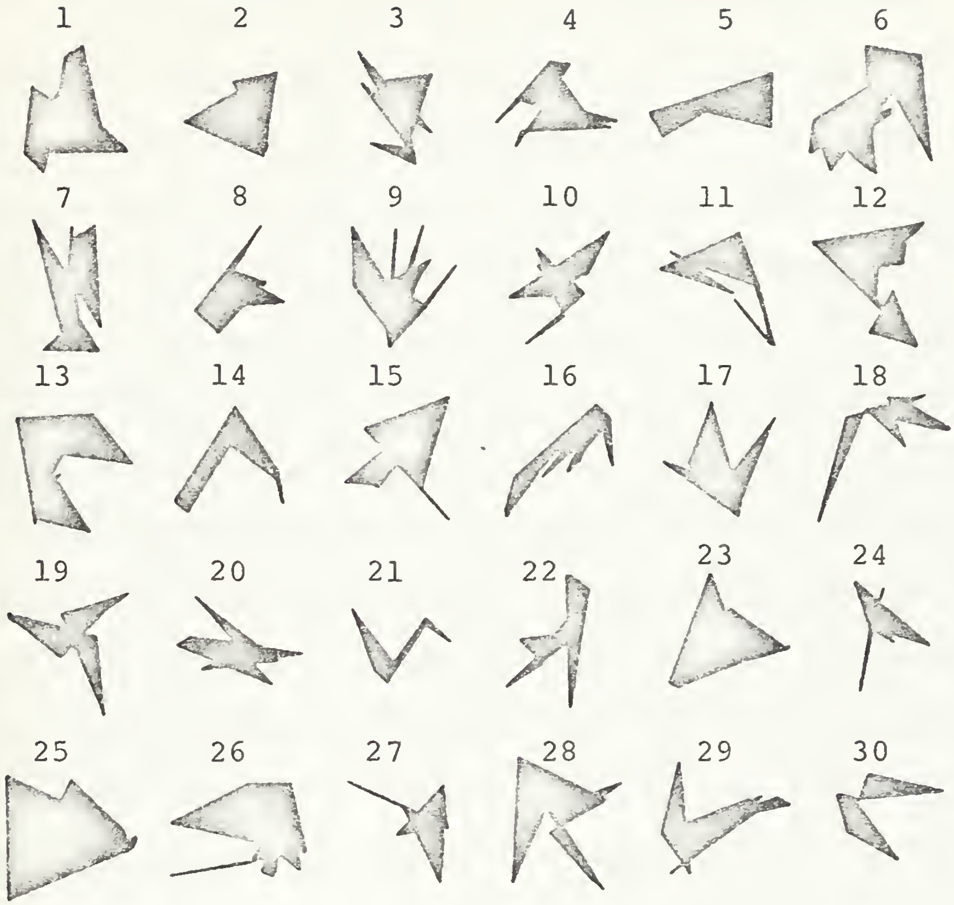

FIGURE 1. Shapes selected for use in assembling stimulus lists.

(From Arnoult, 1956)

4. Experiment I

The 30 stimulus polygons were arranged in pairs. All possible pairs were constructed under the constraint that an item would not be paired with itself. Left-right order within a given pair was not considered. This resulted in the assembly of (30 X 29)/2 = 435 different pairings. These pairs were then arranged in three columns on sheets. Three separate booklets, each containing 145 pairs, were constructed and distributed to 60 graduate students at the Naval Postgraduate School. Each subject received a single booklet selected at random from the three, and was asked to perform two separate tasks--selection of one item from each pair and rating of the degree of similarity seen between the items of each pair. Subjects were told that one item in each pair had been arbitrarily designated as "correct," i.e., the desired response, and were asked to designate that item which they thought to be the "correct" response. This selection was to be made with the knowledge that designation of the "correct" response was made completely arbitrarily.

Subjects were cautioned to make their choices solely on the basis of a given pair alone, and without regard to previous selections. This exercise was intended to simulate as closely as possible the condition of facing a stimulus pair in a forced-choice situation with no prior knowledge of the correct item in the pair.

Subjects then went through the list a second time, rating each pair as to whether the two items in each appeared

to be very similar, slightly similar, or dissimilar. Each pair was then assigned a similarity factor of one, two, or three, respectively. Full instructions for both tasks, as printed on the booklet covers, are presented in Appendix A.

The choice preferences of the 60 subjects (20 for each set of 145 pairs) were translated into percentages and cast into a matrix (Table 1). In addition, averages of similarity ratings given for each pair were computed and cast into the same matrix format (Table 2). Thus pairwise estimates of choice preference and item similarity were obtained and placed in usable form.

5. <u>Construction of Test Stimulus Lists</u>

A subgroup of pairs was selected from the original 435 that had been rated. These pairs were singled out on the basis of choice preference. Subjects making choices within these pairs had displayed no significant preference, on the average, for either item in each pair (selections were distributed either 50%-50% or 45%-55% between each). This subgroup was then used to construct stimulus lists for DL testing. Since no marked preference for a given item in a pair had been demonstrated, it was felt that the choice probabilities associated with each could be considered to be "equally likely" for the purposes of evaluating the information content of the choice associated with each pair.

Three stimulus lists of six pairs each were constructed from the "equally likely" subgroup of pairs. These lists were assembled under the following constraints:

Pairwise Choice Preferences for Thirty Two-Dimensional Polygons Indicated by Sixty Subjects

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
—	50	45	35	30	40	50	40	40	50	40	75	55	65	45	50	85	55	65	65	35	70	75	65	55	55	30	40	70	50	1
	—	45	50	35	35	70	45	60	45	45	65	40	35	55	45	70	55	60	50	45	40	60	45	60	45	35	50	80	45	2
		—	30	65	45	55	45	30	30	60	30	55	35	45	35	70	25	75	50	50	30	60	15	55	30	20	35	50	30	3
			—	65	40	55	75	75	55	30	55	65	75	55	10	70	50	75	60	70	50	75	45	60	35	55	45	55	65	4
				—	30	40	45	40	35	50	50	35	50	45	40	70	40	65	45	65	55	65	30	55	30	50	50	55	65	5
					—	65	70	50	30	30	60	50	75	45	50	70	50	70	75	55	65	65	50	65	35	65	50	60	65	6
						—	60	45	40	50	35	40	60	40	35	70	35	50	55	50	45	65	50	45	10	35	35	55	65	7
							—	60	25	30	40	25	60	45	35	50	45	50	40	70	40	45	30	30	40	15	40	50	35	8
								—	35	55	55	45	25	60	30	70	35	80	65	75	55	45	35	45	40	40	40	60	50	9
									—	40	40	65	50	55	45	75	35	80	55	60	65	65	35	75	40	50	65	80	65	10
										—	50	75	60	60	45	65	40	60	45	70	70	90	45	40	40	40	45	85	60	11
											—	45	40	50	40	85	45	65	70	65	40	35	65	40	40	40	60	65	60	12
												—	35	55	40	65	70	50	60	65	35	75	20	65	40	50	55	60	60	13
													—	60	40	90	25	50	55	65	60	70	55	45	35	50	40	60	45	14
														—	40	70	45	65	45	50	40	60	25	50	50	50	30	40	40	15
															—	85	45	70	80	60	80	80	50	70	40	60	55	75	70	16
																—	15	35	35	50	30	50	80	45	15	15	35	40	40	17
																	—	80	75	65	80	70	55	50	30	50	35	70	55	18
																		—	25	50	20	45	35	30	65	20	30	55	25	19
																			—	70	30	40	05	45	25	25	40	60	30	20
																				—	40	45	35	30	25	40	25	30	50	21
																					—	50	55	65	45	45	55	50	45	22
																						—	15	05	40	60	30	65	55	23
																							—	60	55	55	65	75	55	24
																								—	60	45	40	80	80	25
																									—	70	40	80	85	26
																										—	70	60	70	27
																											—	65	60	28
																												—	50	29
																													—	30

Note:

1. Decimals omitted.
2. Entries indicate percent of time columnar item was chosen over row item.
3. Item designations refer to numbers assigned in Figure 1.

Averaged Pairwise Similarity Ratings of Thirty Two-Dimensional Polygons as Assigned by Sixty Subjects

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	—																													
2	1.30	—																												
3	2.25	2.05	—																											
4	2.05	2.40	1.55	—																										
5	2.05	2.30	2.65	2.50	—																									
6	2.10	2.50	2.35	1.95	2.35	—																								
7	2.05	2.55	1.45	1.75	2.35	1.70	—																							
8	1.80	2.35	2.35	2.15	2.20	2.55	2.50	—																						
9	1.75	2.60	1.80	1.85	2.90	2.50	2.25	2.15	—																					
10	2.35	2.55	2.00	1.95	2.65	2.65	2.35	2.10	1.65	—																				
11	2.50	2.20	2.00	2.20	2.15	2.55	2.40	2.45	2.55	2.20	—																			
12	2.20	2.00	2.15	2.30	2.50	1.95	2.15	2.75	2.40	2.20	2.25	—																		
13	2.40	2.15	2.10	2.45	2.30	1.85	2.55	2.55	2.55	2.55	2.35	1.80	—																	
14	2.55	2.50	2.40	2.70	2.45	2.65	2.40	2.25	2.55	2.35	2.10	2.35	1.35	—																
15	2.05	1.95	1.85	2.35	2.45	2.65	2.55	2.25	2.55	2.35	2.30	2.10	2.50	2.05	—															
16	2.80	2.80	1.70	2.40	2.15	2.30	2.35	2.60	2.00	2.15	2.05	2.45	2.70	1.70	2.65	—														
17	2.10	2.15	1.60	1.60	2.65	2.60	2.10	2.35	1.55	2.35	2.25	2.35	2.05	2.30	1.80	2.40	—													
18	2.85	2.80	1.90	2.60	2.45	2.75	2.80	2.50	2.25	2.10	2.10	2.85	2.65	2.10	2.55	1.55	1.50	—												
19	2.60	2.75	2.35	2.45	2.70	2.70	2.20	2.45	2.50	1.85	2.05	2.35	2.65	2.15	2.60	2.50	2.05	1.90	—											
20	2.45	2.70	1.90	2.45	2.55	2.70	2.10	2.00	1.85	1.90	2.10	2.80	2.75	2.50	2.55	2.40	1.65	2.35	1.80	—										
21	2.75	2.75	2.15	2.45	2.70	2.65	2.50	1.90	2.45	2.50	2.70	2.65	2.70	1.95	2.50	2.40	2.40	1.65	1.95	1.70	—									
22	2.55	2.60	2.20	2.25	2.65	2.45	2.15	2.05	2.40	1.85	2.45	2.50	2.65	2.10	2.20	2.30	2.25	2.10	1.60	1.40	2.30	—								
23	1.45	1.10	2.25	2.35	2.35	2.50	2.70	2.65	2.20	2.55	2.20	2.00	2.30	2.45	1.60	2.75	1.80	3.00	2.50	2.80	2.85	2.90	—							
24	2.60	2.75	2.20	2.35	2.65	2.55	2.10	1.95	2.35	2.00	2.25	2.70	2.90	2.05	2.50	2.20	2.30	1.40	1.90	1.80	1.75	1.50	2.65	—						
25	1.85	1.10	2.55	2.30	2.05	2.35	2.40	2.85	2.35	2.80	2.50	2.15	2.25	2.75	2.15	2.85	2.25	2.95	2.80	2.80	2.85	2.90	1.40	2.90	—					
26	1.75	2.20	2.35	2.15	2.50	1.90	2.90	2.15	2.50	2.40	2.45	2.35	2.10	2.60	2.50	2.60	1.80	2.65	2.55	2.25	2.85	2.85	1.75	2.55	2.00	—				
27	2.75	2.65	1.60	2.25	2.75	2.75	2.60	1.90	2.55	1.90	2.10	2.55	2.75	2.35	1.90	2.35	2.10	2.00	2.30	2.00	1.90	1.75	2.45	1.40	2.55	1.85	—			
28	2.45	1.85	2.10	2.10	2.35	1.70	2.25	2.70	2.05	2.15	2.10	1.65	2.15	2.95	1.60	2.70	2.00	2.60	2.20	2.35	2.70	2.05	2.10	2.50	1.60	2.05	2.75	—		
29	2.30	2.30	2.00	2.20	2.45	2.45	1.75	1.50	1.65	2.00	2.20	2.50	2.30	1.50	2.70	2.30	2.05	1.90	1.90	1.75	1.70	1.95	2.15	1.95	2.65	2.30	2.20	2.05	—	
30	2.50	2.40	1.85	2.45	2.45	2.35	2.45	2.35	2.40	2.40	1.95	2.35	2.15	1.60	2.95	2.45	1.95	1.60	1.65	1.75	1.60	1.90	2.75	2.05	2.75	2.70	2.15	2.50	1.65	—

Note:

1. Entries are averages of subjective evaluations of similarity between two items in a pair in accordance with the following scale:

   1 = quite similar
   2 = somewhat similar
   3 = dissimilar

2. Item designations refer to numbers assigned in Figure 1.

<u>List I</u>: Figures in each pair were as dissimilar as possible. In addition, all figures in the entire list were as dissimilar as possible. (Within-pair similarity factors--from Table 2--were at least 2.50, averaging 2.60, while between-pair factors were no less than 1.75, averaging 1.98.)

<u>List II</u>: Figures in each pair were as similar as possible, but dissimilarity between pairs was maintained. (Within-pair rating factor was no greater than 1.95, averaging 1.58; the between-pair factors were no less than 1.90, averaging 2.20.)

<u>List III</u>: Figures were as similar as possible, both within each pair and between other figures in the list. (Within pair similarity factor was no more than 1.90, averaging 1.73; between-pair factor was no greater than 2.30, averaging 1.92.)

These lists are presented in Figures 2, 3, and 4, respectively.

As can be seen, the lists were constructed in order to present discrimination tasks of increasing difficulty. Stimulus items in List I were chosen to be as distinguishable as possible, minimizing intra- and interpair confusion. Similarity within pairs was added in List II, but each pair was

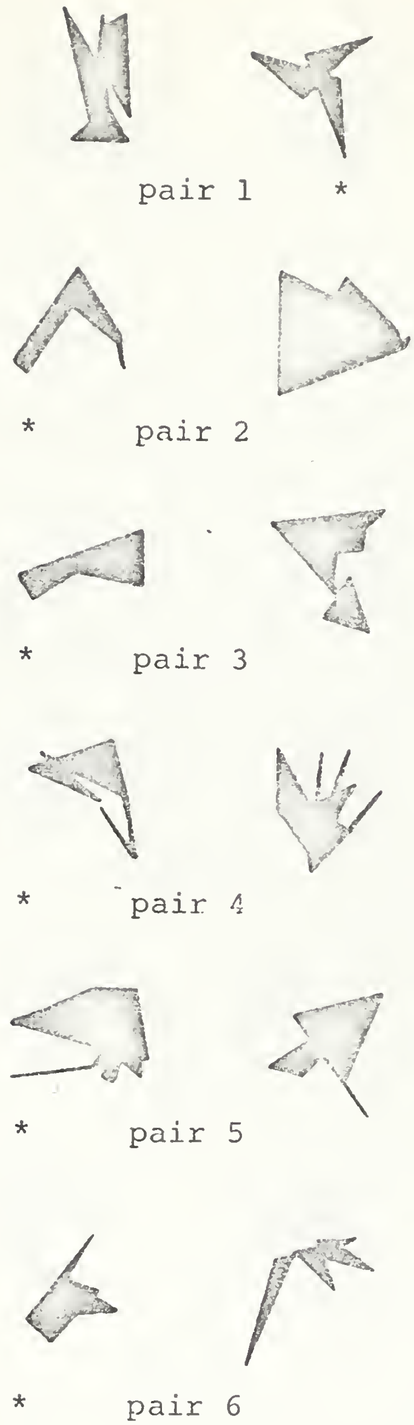

FIGURE 2. Stimulus List I.
(Least similarity within and between pairs)
*Indicates "correct" shape.

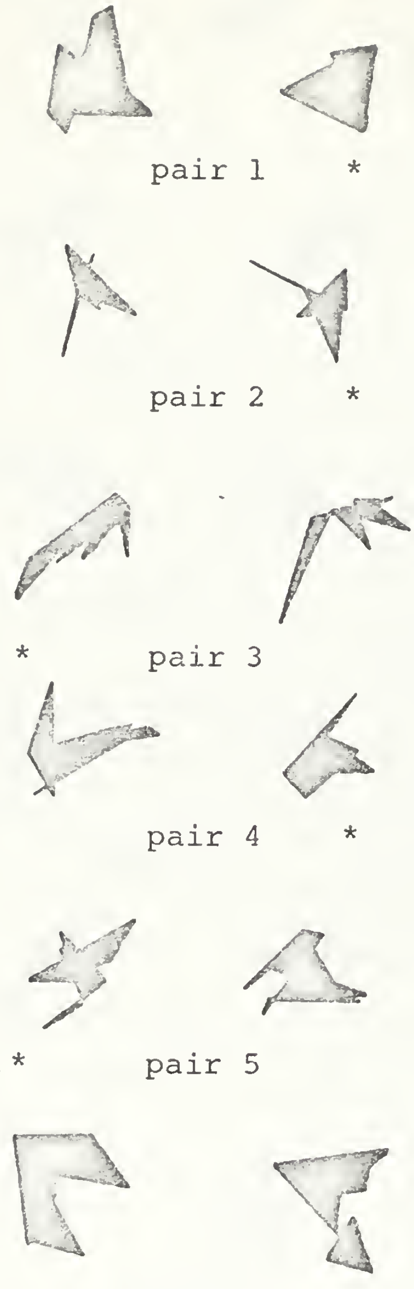

FIGURE 3. Stimulus List II.

(Maximum similarity within pairs; minimum similarity between pairs.)

*Indicates "correct" shape.

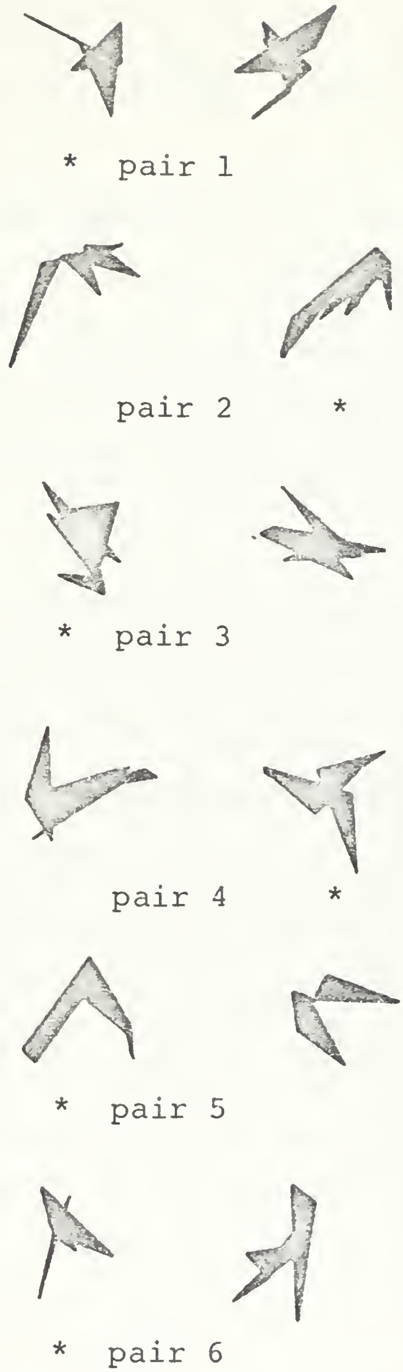

FIGURE 4. Stimulus List III.
(Maximum similarity both within and between pairs.)
*Indicates "correct" shape.

kept as distinguishable as possible from other pairs in the list. Similarity was extended to cover all items in List III.

When lists of six pairs each had been completed, test stimulus lists of 60 pairs were assembled. Each test list consisted of 10 repetitions of each of the six pairs of Lists I, II, and III. Construction of the 60-pair lists was performed on a pseudo-ramdom basis with the following restrictions:

(a) Lists were subdivided into 10 replicates, each of which contained the basic list of six pairs. Order within these replicates was pseudo-random in order to give the appearance of overall randomness but still maintain discrete groupings of stimuli.

(b) Left-right order within the pairs was varied in a pseudo-random fashion as well, but was such that a given item was seen on the right five times and on the left five times in order to preclude positional cues.

(c) At least one different pair was presented before a given pair was repeated.

(d) Polygons were not rotated or reversed, but were presented "upright" at all times (Arnoult, 1954).

The resultant sets of 60 pairs are listed in Table 3.

Table 3

Stimulus Set Ordering for all Three Stimulus Lists

Replicate	List I	List II	List III
1	2:1:5:6:3:4	6:4:5:1:3:2	6:3:5:4:1:2
2	1:2:6:4:3:5	1:4:5:3:6:2	5:3:4:2:1:6
3	6:2:5:3:1:4	5:1:4:2:3:6	4:3:6:2:5:1
4	2:3:6:4:1:5	1:2:3:4:5:6	2:6:4:3:5:1
5	4:5:2:3:6:1	4:3:1:6:2:5	3:6:5:4:1:2
6	5:2:6:4:3:1	6:3:4:5:2:1	6:5:2:1:4:3
7	4:5:6:3:2:1	4:2:1:5:6:3	1:4:5:2:3:6
8	3:6:1:2:5:4	2:5:4:3:1:6	4:1:3:5:6:2
9	6:1:2:4:5:3	3:5:2:6:4:1	5:3:6:1:4:2
10	2:4:5:3:1:6	2:5:1:4:6:3	1:4:6:3:2:5

Note. Item designations refer to numbers assigned to stimulus pairs in Figures 2, 3, and 4.

Thus each test subject could be presented a total of 60 pairs of stimuli. Pairs appeared in no apparent order, and the correct response was not always on either the right or left side; sugjects were forced to learn the correct response in each pair solely on the basis of recognition of the items within that pair alone.

C. TEST APPARATUS

Test apparatus was designed to provide maximum flexibility in test administration. The apparatus array used in administering the test is diagramed in Figure 5. Critical units of the presentation and response equipment were secured in place throughout the course of test administration. Distance from the subject (edge of table) to the viewing screen was 42.5 inches (107.95 cm); reinforcement lights were located 8.5 inches (21.59 cm) in front of the screen. Stimulus pairs occupied an area on the screen approximately 6 inches (15.24 cm) high by 9 inches (22.86 cm) wide.

Stimulus pairs were mounted on 35mm slides, one pair to a slide. Since each list was presented a total of 10 times, the 60 slides required for each list were placed in a carousel. Stimuli were rear projected onto a Kodak shadow-box screen using a Kodak Ektographic Carousel slide projector, Model B-2. A neutral light-reduction filter (Kodak Wratten gelatin filter, no. 96 ND 0.50), rated to reduce light transmission by 50 percent, was fixed over the projector lens to reduce excessive glare on the screen.

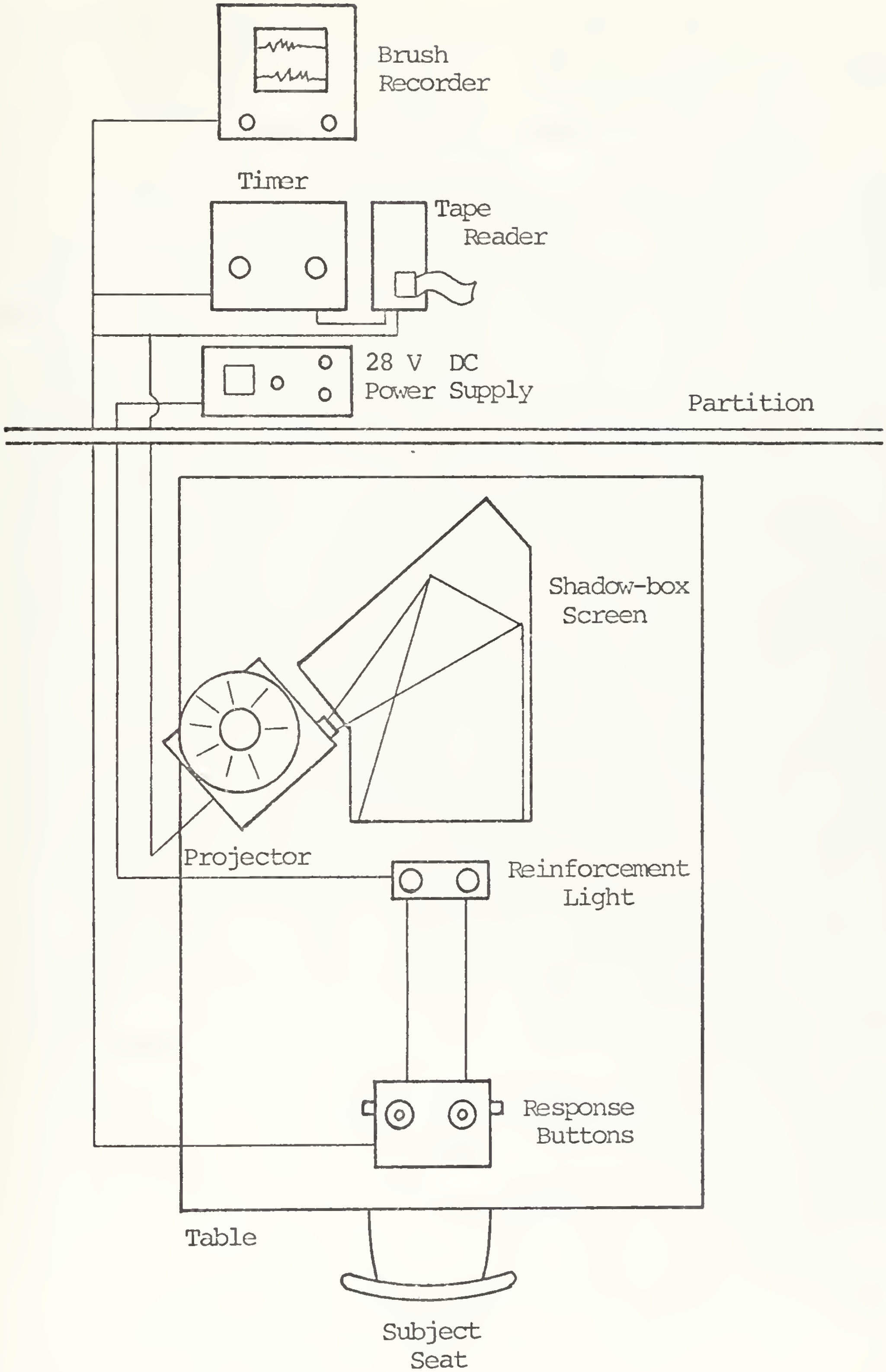

FIGURE 5. Layout of Test Equipment

A modified Ohr-tronics eight-channel paper-tape reader was used to control the reinforcement lights (described below) so that only correct responses would receive reinforcement. Wiring was accomplished so that the pulse used to advance the slide projector to the next stimulus pair also advanced the tape reader. Tapes were punched to co-ordinate with the ordering of the stimulus list in use.

The apparatus was designed to permit a machine- or self-paced mode of presentation. Stimulus presentation rate in the machine-paced mode was controlled by a Lafayette Model 5004B timer. The timer was set to provide an actuating pulse to both projector and tape reader simultaneously every 4.0 seconds. The time required for the slide projector to cycle from a presented slide to the next slide was found to be 1.0 sec. Since the projection screen was blank during this cycle time, the stimulus pairs were visible for only 3.0 sec before the timer initiated the next sequence. Thus an IPR of 1/3 bits per sec was accomplished with 1.0 sec between stimuli.

Stimulus presentation during the self-paced mode was controlled by either of two identical buttons located on the sides of the response box. Pressing either of these buttons initiated the electrical pulse that advanced the slide projector and tape reader. (These buttons were inactivated during the machine-paced mode to preclude accidental disruption of the stimulus presentation rate.)

Two identical buttons fixed on top of the response box were used to designate choices. Correct responses were

reinforced by one of a pair of 2.5 watt lights placed on a small box directly in front of the viewing screen. Incorrect responses received no reinforcement. Responses, regardless of reinforcement, were recorded on a two-channel Clevite brush recorder, Model Mark 220. The tapes thus obtained could be used to confirm observed responses, and in the self-paced mode to measure inter-response time and total test time.

Twenty-eight volt DC current to power the tape reader and reinforcement lights was obtained from a Power Designs, Inc., Model 3650-S DC Power Supply.

Simplified schematic representation of the apparatus wiring is shown in Figures 6 and 7 for the self-paced and machine-paced phases, respectively.

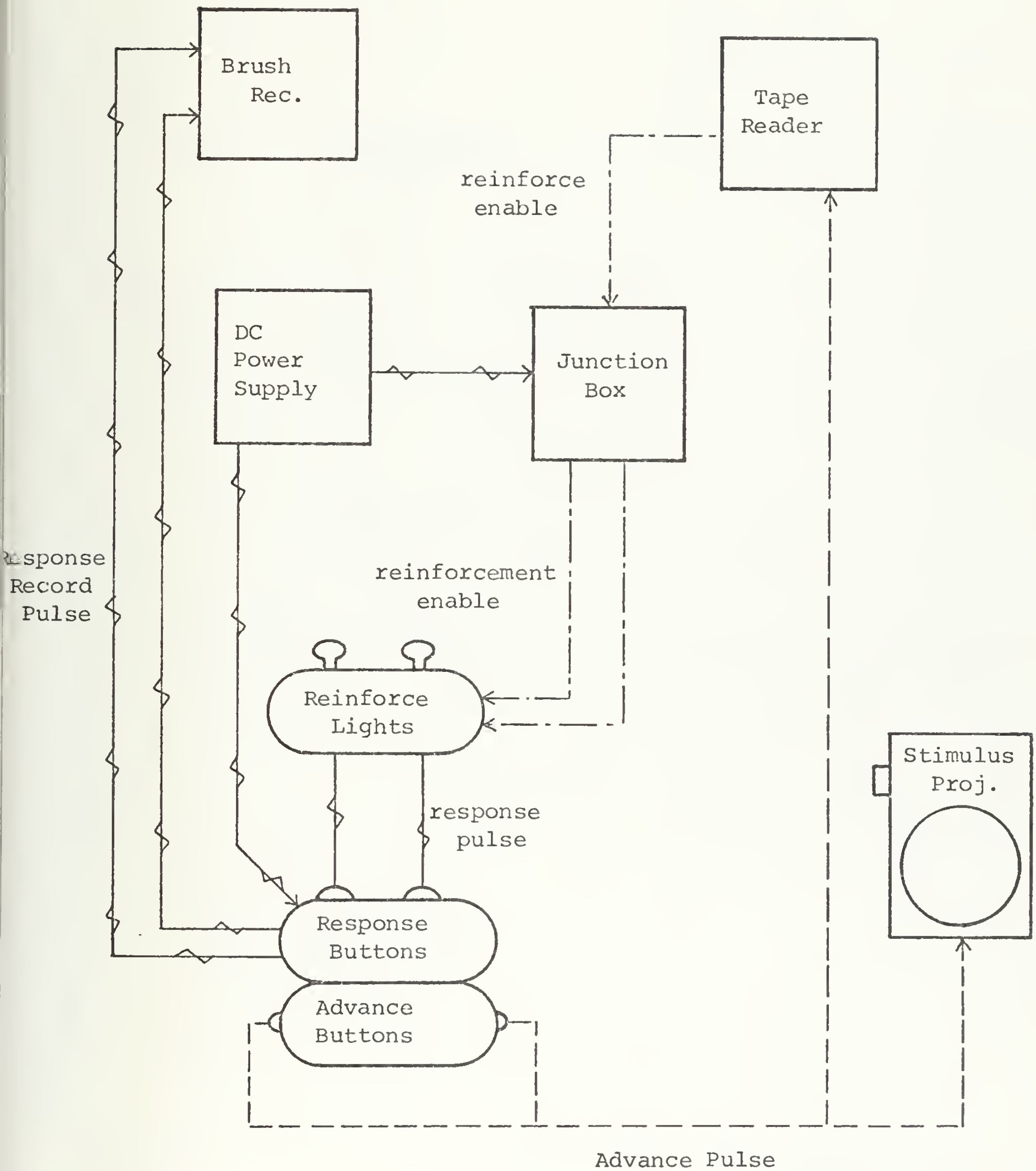

FIGURE 6. Block Diagram of Test Equipment in Self-Paced Mode.

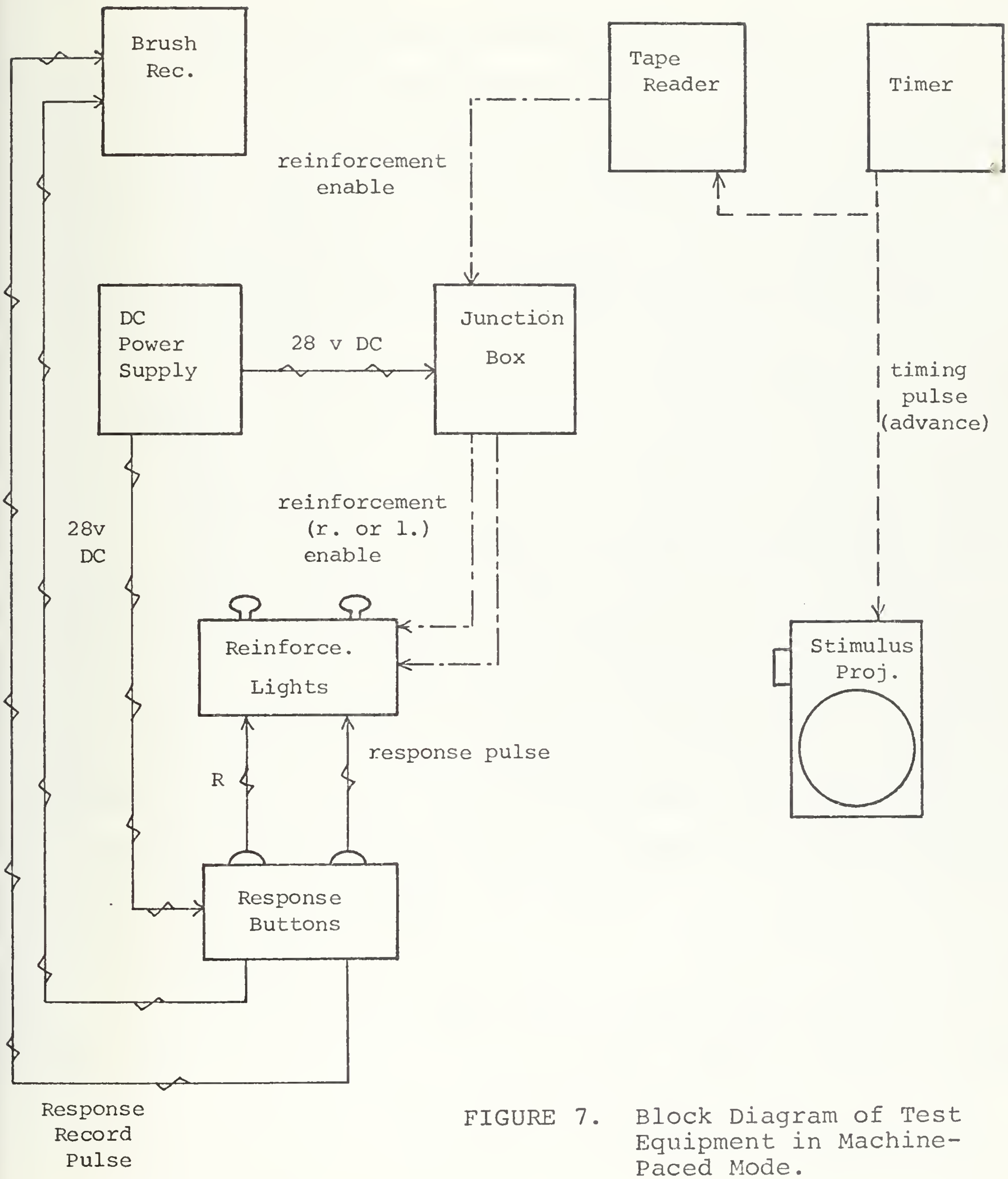

FIGURE 7. Block Diagram of Test Equipment in Machine-Paced Mode.

III. EXPERIMENT TWO: TEST TRIAL

In order to evaluate the characteristics of the constructed test under conditions as close to operational as possible, and also to investigate the appropriateness of the various test parameters (IPR, list length and compositon, etc.), it was decided to administer the test to as many subjects as possible during a five-week period in which they would be available.

A. METHOD

1. Facilities

Testing was conducted in November and December, 1973, at the Naval Training Center (NTC), San Diego, California. All testing was performed in an isolated room at the Personnel Testing and Classification Center located on board NTC. Since activity was planned for both morning and afternoon periods, windows in the testing room were covered with opaque material to reduce anticipated glare from sunlight and to achieve uniform lighting conditions in the room.

2. Subjects

Subjects tested were 160 male U.S. Navy recruits at NTC. Ages ranged from 17 to 26 years, with the average being 19 years. Average stated schooling level for the group was 12th grade (11.78). Schooling level within the nonwhite subgroup was slightly higher (12.2 years) than the group average. Nonwhite subjects were predominantly Negro, although the

sample contained Oriental, Maylay (Filipino), and Mexican-American recruits. Subjects were assigned to the various test conditions in order of appearance.

3. Test Design

The experiment was conducted in four major phases. Forty-four subjects were given the test using self-pacing to control the stimulus presentation rate. Stimulus List I was used throughout the self-paced phase. The remaining three phases were machine-paced to present the stimulus pairs at a constant rate of one each 4 secs. A one-second inter-stimulus time (cycle time of the projector) thus gave a 1/3 bit-per-second IPR. In the three machine-paced phases, 43, 40, and 33 subjects were tested using Stimulus Lists I, II, and III, respectively. Tabular representation of this test design is shown in Table 4.

Table 4

Test Design

Test Group	Subjects (White; Nonwhite)	Pacing	Stimulus List
1	44 (31; 13)	Self	I
2	43 (30; 13)	Machine	I
3	40 (31; 9)	Machine	II
4	33 (29; 4)	Machine	III

4. Procedure

Subjects were brought into the testing room in groups of not more than six. The apparatus was displayed, and the experimental nature of the testing explained briefly prior to issuing the verbal instructions contained in Appendix B. Instructions emphasized the nature of the stimuli, what was required of the subject in the way of response, and the operation of the apparatus itself. Subjects were then given the opportunity to ask questions about the test and procedure, and to decline participation if they so desired. They were then asked to wait outside the room and were brought in for testing one by one. The instructions for the test were then reviewed with each individual as he was seated at the response box prior to commencement of the experiment.

Stimulus pairs were then presented one by one on the viewing screen in the order given in Table 3 for his test condition. Each list of six pairs was presented in 10 consecutive trials with no break between lists. As a subject selected the figure in each pair that he thought was correct, he pressed the corresponding (right or left) response button in front of him. Correct responses were reinforced by a small light in front of the view screen, while incorrect responses received no reinforcement. The IPR was determined as described above.

As testing was in progress, the experimenter stood behind the subject and recorded his responses on an answer sheet. Responses were also recorded electrically on a

two-channel Brush recorder. Upon completion of the test, the subject was cautioned not to discuss anything he had seen or done in the test with those who had not yet been tested. This request was repeated to the entire group after all had been through the test.

Performances by six of the original 160 subjects were discarded. Improper operation of the self-pacing buttons that put the tape reader out of phase with the projector was cause for rejection of three performances. Another subject in the first (self-paced) group was unable to follow instructions. Timer malfunction caused two performances in the first machine-paced group to be eliminated.

Seventeen other subjects' performances were not used in the data analysis because their BTB scores and/or demographic data could not be retrieved from computerized records. As a result of these subject losses, the 137 remaining subjects (white and nonwhite) were distributed as follows: Group 1 (24, 11); Group 2 (25, 12); Group 3 (28, 8); and Group 4 (30, 3).

IV. RESULTS

Individual performances in the test, in the form of number of correct choices made per trial per unit of time, were computed to arrive at the test measure of effectiveness, Information Processing Rate (IPR). Dimensions of IPR were bits of information correctly processed per second. Performances in the first trial were not used, since responses in the initial trial were dependent wholly upon chance, and as such were not indicative of learning ability.

The number correct in each trial was divided by the amount of time the stimuli were presented to the subject. (In the machine-paced mode, this was a constant 3 sec. per pair. Scores for the self-paced group were scaled to individual rates.) In both situations, the 1 sec. cycle time (interstimulus time) of the slide projector was not included in computing IPR. The resultant trial IPR scores were grouped into three blocks of three consecutive trials each. These figures are listed in Table 5. Rates of processing information are seen to generally increase over blocks of trials for all groups. (The single exception is the nonwhite subset of test group Four, where performance declines very slightly over trials. This group contained three subjects.) Overall performances by all groups were quite similar, despite differences in pacing mode and stimulus similarity between groups. Overall performance by the nonwhites in test group One

(self-paced) exceeded that of the whites; the reverse was true for the three machine-paced groups. Figures 8 and 9 depict aspects of these situations.

The results listed in Table 5 were subjected to an analysis of variance using a three-way design compensating for unequal cell populations by test group, racial group, and blocks of trials as described by Kirk (1968). This design utilizes the harmonic mean to estimate within-cell degrees of freedom. The results of this analysis are presented in Table 6.

As can be seen in Table 6, significant effects were noted between racial groups and among blocks of trials. Analyses of variance were also conducted using a one-way, repeated measures design in order to determine the contribution of between-subjects variability to the overall error term of Table 6. These analyses were run for four groups, established on the basis of race and pacing mode, and the results are listed in Table 7. Significant between-subject and between-blocks effects are seen in all groups.

Table 8 shows the results of an analysis of variance conducted using a two-way design for unequal cell frequencies (Winer, 1962) on the four groups established in Table 7. With subjects grouped in this manner, no significant effect is noted as a result of pacing mode or racial grouping. Because performances by nonwhites were seen to exceed those of whites in the self-paced mode and yet lag behind in the machine-paced groups, an analysis of variance was conducted using only the machine-paced groups. The results of this analysis, setting race against stimulus sets, and using the same design to allow

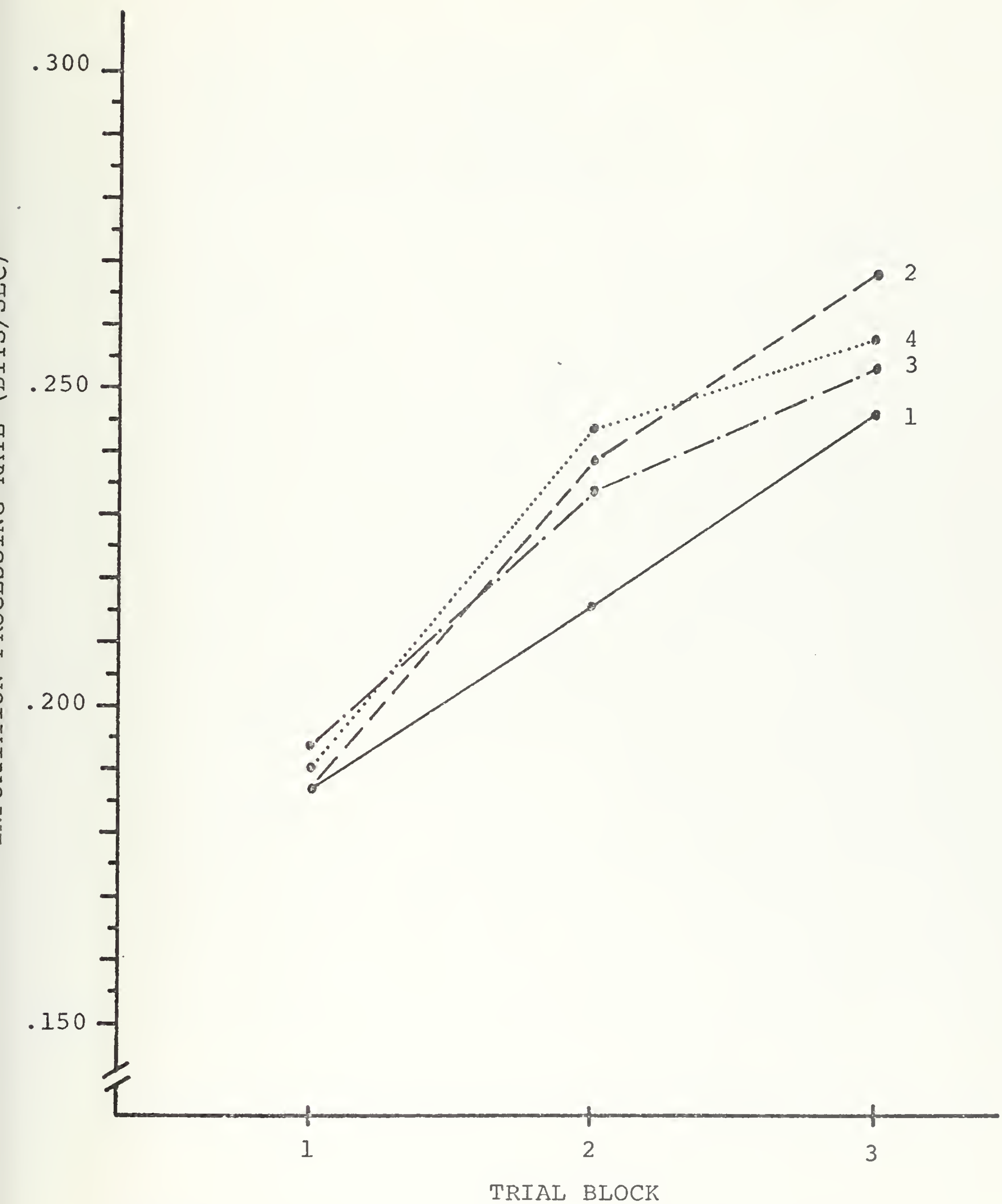

FIGURE 8. Information Processing Rate by Test Group and Blocks of Trials.

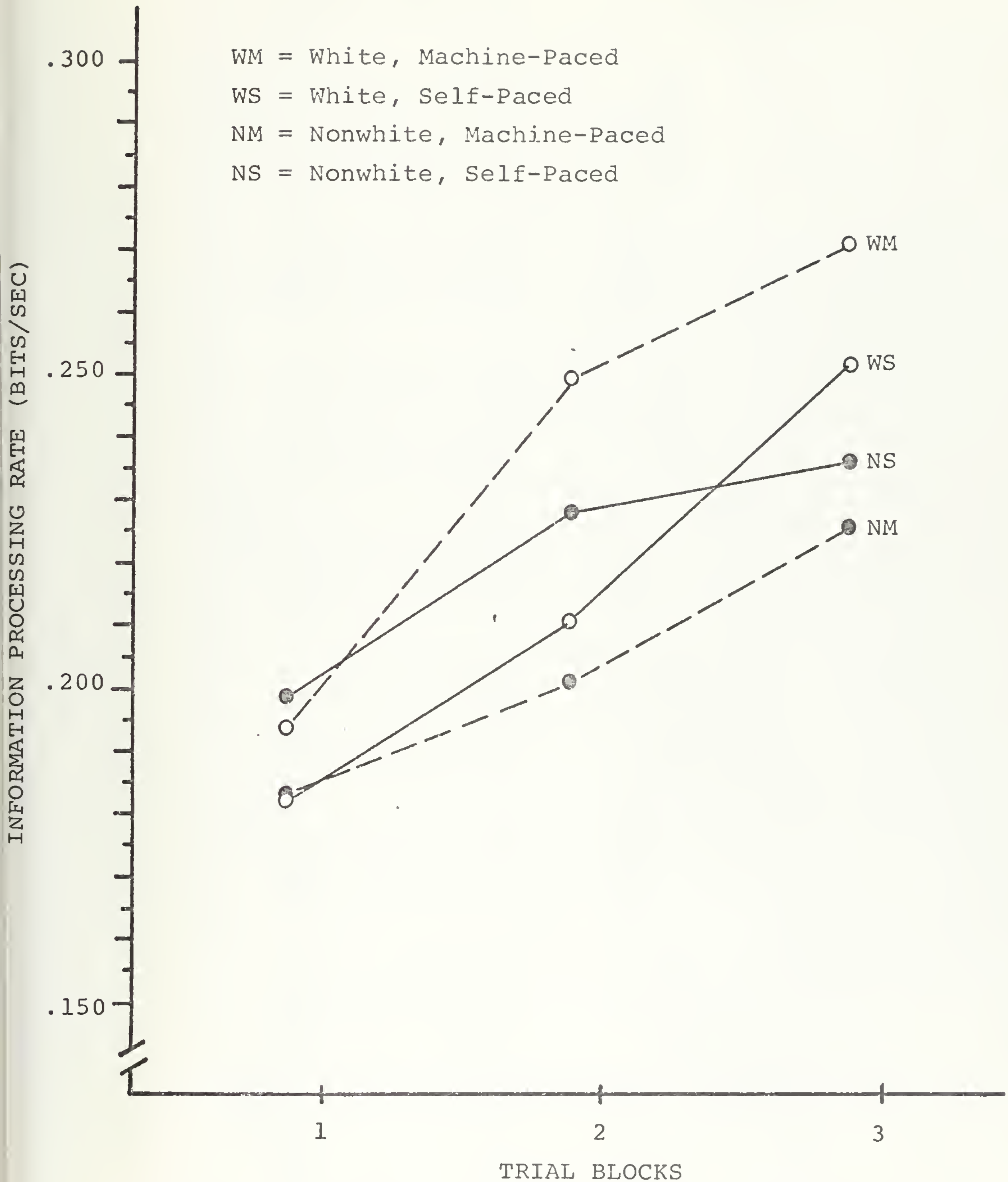

FIGURE 9. Information Processing Rate by Racial Group, Pacing Mode, and Blocks of Trials.

Table 5

Information Processing Rate in Multiple Discrimination Learning by Test Group, Blocks of Trials, and Racial Group

Test Group	Block 1		Block 2		Block 3		Totals		
	White	Nonwhite	White	Nonwhite	White	Nonwhite	White	Nonwhite	Combined
1	181.416	198.000	210.666	227.818	251.958	235.636	214.291	220.484	216.580
2	202.520	156.916	261.640	193.916	284.000	237.333	249.386	196.055	232.089
3	190.071	205.500	240.892	207.750	266.607	212.500	233.713	228.129	232.472
4	187.418	222.000	247.074	215.666	262.814	214.333	231.962	230.533	231.819
Totals									
White	190.009	—	240.509	—	267.461	—	232.659	—	—
Nonwhite	—	187.382	—	210.058	—	228.999	—	208.813	—
Combined	189.361		233.000		257.984				226.783

Note. Entries are bits/sec X 10^3.

Table 6

Analysis of Variance of Overall Performance by Test Group, Racial Group, and Blocks of Trials

Term	*df*	*SS*	*MS*	*F*	*p*
Total	243	1,829,659.50	—	—	—
Test Group (T)	3	5,082.50	1,694.10	0.230	n.s.
Racial Group (R)	1	31,511.00	31,511.00	4.288	<.05
Trial Block (B)	2	117,910.00	58,955.00	8.023	<.001
T X R	3	24,396.00	8,131.90	1.106	n.s.
T X B	6	10,165.00	1,694.10	0.230	n.s.
R X B	2	21,346.00	10,673.00	1.452	n.s.
T X R X B	6	13,214.00	2,202.30	0.299	n.s.
Error	220	1,616,200.00	7,347.60	—	—

Table 7

Analysis of Variance of Overall Performance by Subject and Blocks of Trials

Term	df	SS	MS	F	p
Nonwhite, Self-Paced					
Total	32	495,786.24	15,493.32	—	—
Subject	10	468,473.58	46,847.36	50.279	.001
Block #	2	8,678.06	4,339.03	4.657	.025
Error	20	18,634.61	931.73	—	—
White, Self-Paced					
Total	71	411,441.65	5,794.95	—	—
Subject	23	313,704.99	13,639.35	16.756	.001
Block #	2	60,293.53	30,146.76	37.036	.001
Error	46	37,443.14	813.98	—	—
Nonwhite, Mach-Paced					
Total	68	274,322.29	4,034.15	—	—
Subject	22	168,634.96	7,665.23	4.024	.010
Block #	2	21,878.46	10,939.23	5.743	.001
Error	44	83,808.87	1,904.75	—	—
White, Mach-Paced					
Total	236	814,783.30	3,452.47	—	—
Subject	78	407,124.63	5,219.47	41.15	.001
Block #	2	252,426.43	126,213.22	126.84	.001
Error	156	155,232.24	126.84	—	—

Table 8

Analysis of Variance of Overall Performance by Racial Group and Pacing Method

Term	df	SS	MS	F	p
Total	137	465,094.994	—	—	—
Racial Grp (R)	1	4,417.475	4,417.475	1.310	n.s.
Pacing Mode (P)	1	242.501	242.501	0.071	n.s.
R X P	1	8,772.961	8,772.961	2.602	n.s.
Error	134	451,662.057	3,370.612	—	—

Table 9

Analysis of Variance of Overall Performance by Racial Group and Stimulus Set (Machine - Paced Only)

Term	df	SS	MS	F	p
Total	102	5,316.928	—	—	—
Racial Grp (R)	1	342.169	342.169	6.810	.020
Stimulus Set (S)	2	4.758	2.379	0.047	n.s.
R X S	2	96.117	48.058	0.956	n.s.
Error	97	4,873.884	50.246	—	—

for unequal sets, are seen in Table 9. In this case, race is seen to be a significant factor, while no apparent difference is seen between the performances on each stimulus set, despite the graded difficulty (similarity) of each list.

Internal reliability of the test itself was investigated using a split-half design for each test group and each racial group as well as for overall performances. Processing rates were compared for trials 4, 6, and 8 against those of trials 5, 7, and 9. In addition, scores on the latter group of trials were compared with those obtained on trials 6, 8, and 10. The former comparison will be referred to as "low trials" and the latter, as "high trials."

Correlation coefficients thus obtained were used in the Spearman-Brown formula for split-half correlations. Both the raw coefficients and the Spearman-Brown coefficients are listed in Table 10. A majority of the coefficients are seen to be statistically significant.

The relationship between scores on the experimental test and the traditional methods of measuring Navy recruit potential was investigated using the test subjects' scores on the Navy General Classification Test (GCT), a major portion of the standard Basic Test Battery (BTB). The basis for the GCT lies in verbal ability, since the test consists of sentence completions and verbal analogies. Test scores are scaled on a normalized distribution with a mean of 50 and a standard deviation of 10. Performance on the Arithmetic Reasoning Test (ARI) is often combined with GCT scores to obtain a rough

Table 10

Split-Half Reliability Coefficients

Group		Low Trials (468 vs 579)		High Trials (579 vs 6810)		Totals	
		r(raw)	r(S-B)	r(raw)	r(S-B)	Low	High
1	White	.767	.868**	.713	.832**	.865**	.872**
	Nonwhite	.756	.861**	.864	.927**		
2	White	.800	.889**	.865	.928**	.871**	.921**
	Nonwhite	.700	.824**	.826	.905**		
3	White	.615	.762**	.632	.775**	.722**	.759**
	Nonwhite	.367	.537	.535	.697*		
4	White	.674	.805**	.664	.798**	.802**	.794**
	Nonwhite	.637	.778	.610	.758		
	Totals						
	White		.835**		.843**		
	Nonwhite		.788**		.873**		
	Combined		.824**		.851**	.838**	

*Significant at $p < .05$.

**Significant at $p < .01$.

"multiple" used in determining Navy technical school eligibility and aptitude.

Pearson product-moment correlations were computed between test scores and GCT scores obtained from individual service files. These correlations were determined for racial subgroups of subjects falling below and above the GCT mean score of 50, for both racial groups *in toto*, and for the entire sample. These figures are seen in Table 11. Significant values of the correlation coefficient are noted only in the white group as a whole and for the entire sample. Nonwhite test scores did not correlate significantly with GCT performance.

Table 11

Correlations of Test Performance (IPR) with Navy General Classification Test (GCT) Score

		Group Averages		Correlation Coefficient		
		GCT	IPR	GCT GRP	Race GRP	Total
Nonwhite N=33	Low (<50) N=24	42.67	.208	.316	.213	.270**
	High ($\geq$50) N=9	56.89	.207	.601		
White N=104	Low (<50) N=17	42.18	.207	.253	.223*	
	High ($\geq$50) N=87	59.63	.238	.050		

*Significant at $p < .05$.

**Significant at $p < .01$.

Additional insight into the relationship between general (nonverbal) learning ability and verbal intelligence was obtained through the use of linear regression techniques. Figures obtained for the full sample, as well as the white and nonwhite subgroups are listed in Table 12. A further breakdown of the nonwhite subgroup into self-paced and machine-paced units is displayed in Table 13. These relationships are graphically presented in Figures 10 and 11.

It should be noted that the number of scores used in these investigations involving GCT scores was one less than that used in previous calculations. The GCT score of one nonwhite subject could not be obtained from computerized records. Therefore, the number of nonwhite subjects was reduced from 34 to 33 for these calculations alone.

Table 12

Linear Regression of Test Performance (IPR) Against Navy General Classification Test (GCT) Score

Group	IPR (X)		(GCT (Y)		Correlation		Y-
	Mean	S.D.	Mean	S.D.	Coefficient	Slope	Intercept
White (N=104)	.233	.049	56.76	8.33	.223*	28	47.98
Nonwhite (N=33)	.208	.081	46.55	7.87	.213	21	42.28
Combined (N=137)	.227	.059	54.31	9.30	.270**	42	44.72

*Significant at $p < .05$.

**Significant at $p < .01$.

Table 13

Linear Regression of Test Performance (IPR) Against Navy General Classification Test (GCT) Score by Pacing Mode (Nonwhites Only)

Pacing	IPR (X)		GCT (Y)		Correlation		Y-
	Mean	S.D.	Mean	S.D.	Coefficient	Slope	Intercept
Self (N=11)	.220	.125	44.36	8.55	.154	11	42.04
Machine (N=22)	.201	.051	47.64	7.47	.401	59	35.72
Combined (N=33)	.208	.081	46.55	7.87	.213	21	42.28

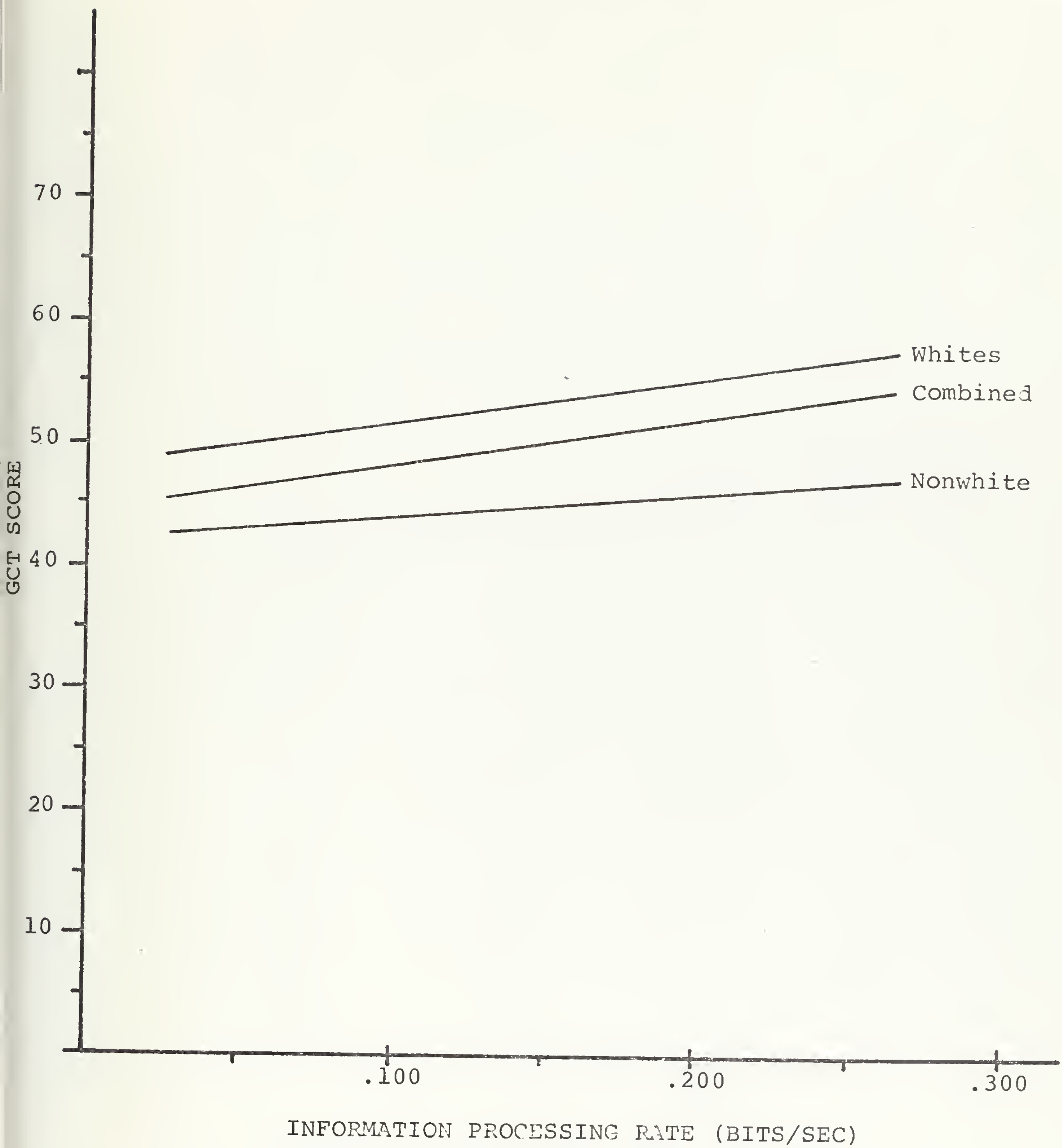

FIGURE 10. Linear Regression of Test Performance (IPR) Against Navy General Classification Test (GCT) Score.

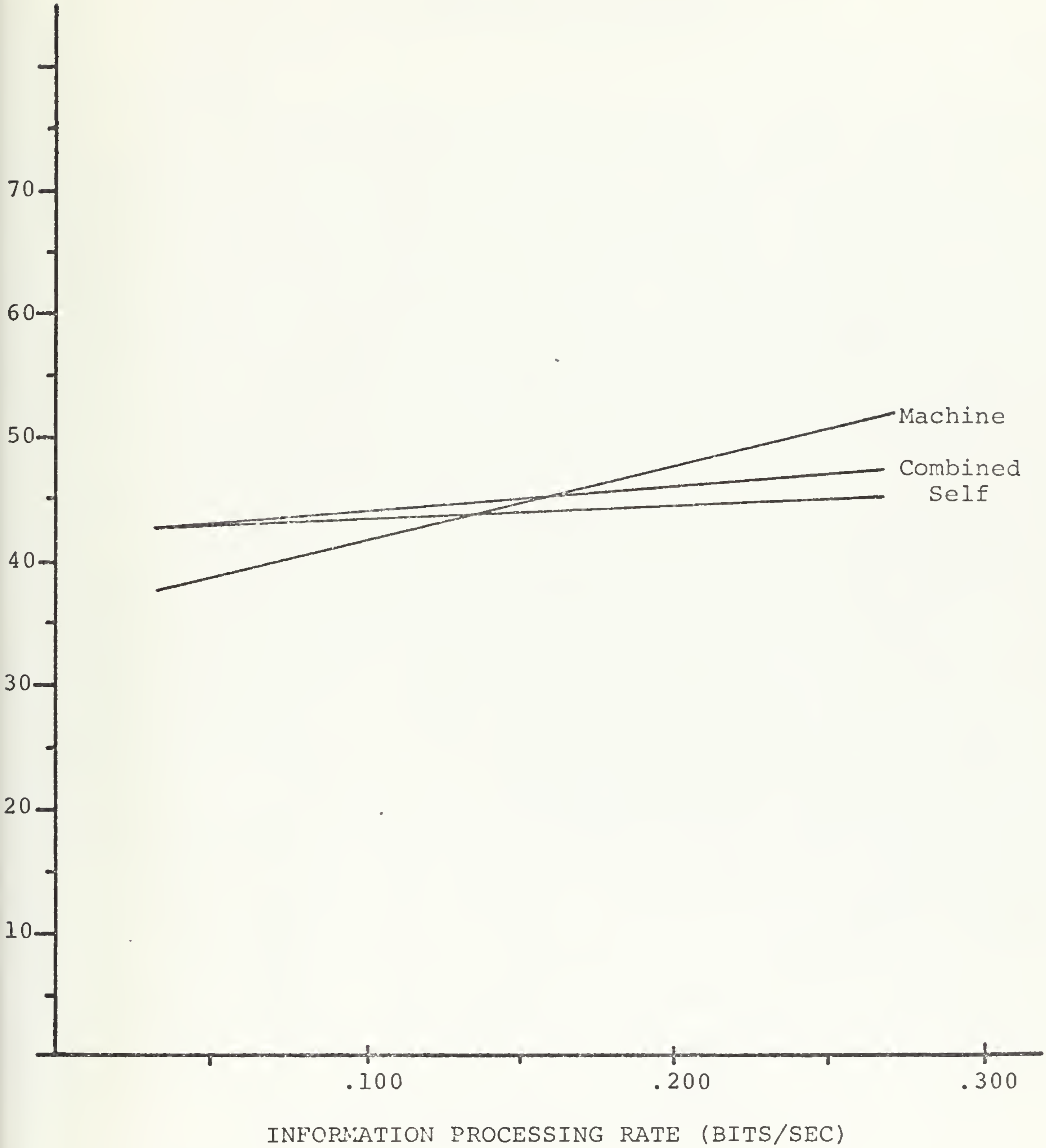

FIGURE 11. Linear Regression of Test Performance (IPR) Against Navy General Classification Test (GCT) Score by Pacing Mode (Nonwhites Only).

V. DISCUSSION OF RESULTS

A. GENERAL

Initial inspection of the experimental results shows that learning did, in fact, take place in the course of the test. In all test and racial groups but one the amount of information processed increased as the trials progressed. The significance of this increase was confirmed by analysis of variance. The only group that did not display this learning effect was considered to be too small ($N = 3$) to provide any sort of conclusive evidence.

The machine-paced mode of information presentation necessarily limited the maximum possible IPR to that of the presentation rate (.333 bits/sec). Comparison of the IPR's obtained during the third block of trials shows that whites attained a maximum of 80 percent of this "perfect learning" rate, while nonwhites reached 69 percent of this quantity. Inspection of the IPR's for the three blocks shows that the greatest marginal learning occurred during the initial trials with the increases on the remaining trials becoming proportionally smaller. This effect corresponds to the classical "learning curve " of grouped data, showing rapid gains early in the testing, with learning rates tapering off as trials are continued. That this effect was present in this case may be due, at least in part, to generalization of stimulus cues between pairs. This factor is suggested in literature on verbal learning tasks (Gibson, 1942, 1959).

At the very least, even though some individuals did accomplish "perfect learning" well before the last trial, the test was not continued until all, or even nearly all, subjects had learned the correct response in all six pairs. Thus the time to perfect learning is not known from these results. Of greater value, however, is the ability to differentiate between individual performances over the same test area. Such differences are, as a rule, more apparent at the intermediate stages than toward the "flatter" end of the learning curve. Thus, stopping the testing at the tenth trial, while a somewhat arbitrary decision when made, was a better point than the perfect learning point, although it may well not have been the optimal point at which to terminate.

Some caution should also be exercised in interpreting the experimental data due to restriction of range in the test sample's measured abilities. Because the testing was accomplished using men who had already been inducted into the Navy, the population from which the sample was drawn represented a pre-selected group, 80 percent of whom were eligible for technical school training upon entry. These recruiting standards ensure that the majority of recruits accepted into the Navy represents at least the 50th percentile of the service-age population in general

B. SELF-PACING vs. MACHINE-PACING

No significant difference in IPR was noted between the self-paced and machine-paced groups. This is notable in light

of the fact that the self-paced group was under no constraint to reach a set or specified rate. These findings tend to support the findings of Arima and Gray (1972), namely that IPR is not affected by presentation rate. An advantage of the self-pacing design is that it allows a superior performer to seek his own level of accomplishment without being limited by a set presentation rate. In the course of testing, the performances of two nonwhites and four whites in the initial group exceeded the machine-paced groups' presentation rate of 1/3 bit per sec. Another effect noted in this case was that nonwhites performed on a par with (actually, somewhat better than) the whites in the self-paced situation, but did significantly poorer when stimuli were presented at a fixed rate. It was in this aspect that the test proved to be "culture-fair."

Self-pacing would appear to be a better choice for this type of test, in that the widest range of test scores is possible. This feature, coupled with the lack of cultural bias noted above, indicates that this would be a useful selection tool in that individual differences would be made more apparent. It is also possible that examination of test-taking strategies in a self-paced situation might give valuable insight into elements of individual personality and motivation. In short, a test with a fixed presentation rate can answer the question, "How much was learned?" A self-paced test can find out, "How much was learned, and <u>how fast</u> was it learned?" This ability <u>per se</u> renders such a test the more useful of the two (Estes, 1974).

Disadvantages to self-pacing are seen to be twofold. For one, there is no theoretical upper limit to the scores. This is not seen as a great problem. The other drawback is seen to be the greater variability apparent in the self-paced scores in comparison to the machine-paced performances. Again, however, this is not seen as a serious problem. In fact, it may be indicative of greater differentiation between individuals accomplished in the course of testing.

C. STIMULUS LISTS

Despite the fact that three stimulus lists were constructed to give graduated degrees of similarity between items, and thus graduated difficulty, performances on the three lists were not significantly different. An implication of this effect is that construction of equivalent stimulus lists is made simple. Explanation of this phenomenon is not quite so easy. A possibility is that, since the lists were all two-choice situations, the basic informational aspects of the choice itself prevailed, namely that the information content of the stimulus lies in the number of alternatives presented, rather than the information content or relative similarity of the figures themselves. It is also possible, however, that the actual discriminations made by the subjects during testing were made on a much finer level than were the original judgments of similarity. Thus, if each figure was seen as distinct from the others at the outset, the confusing effect of intra- and interpair similarity was nullified. On this it is important to note that the original thirty

figures had been selected for their "discriminability" in the study by Arnoult (1956). Thus, even though the three lists had been scaled for similarity, the figures in the lists were apparently not so similar as to make ready discrimination difficult.

It is interesting to note here that several tests of verbal discrimination ability (Arima & Gray, 1972; Baltutis, 1972; Bugarin, 1973) used word (stimulus) lists containing high and low similarity items, but found that this had no effect on IPR. All the words used occurred very frequently in normal language use, and as such were readily distinguished and identified *per se*.

D. RELATIONSHIP OF IPR TO MEASURED INTELLIGENCE

Direct correlation between general learning ability (IPR) and measured verbal intelligence (GCT) was seen to be small for both racial groups. Internal reliability was shown to be strong, and as such was eliminated as a possible reason for this low correlation. In addition, sample scores were inspected to determine if restriction of range had been an unanticipated factor. Table 14 shows, however, that this was not the case. Since the aim of this study was to develop a test of an area not measured by current, verbal-oriented tests, the lack of strong correlation between subjects' performances on the two tests indicates that the instruments do, indeed, measure different abilities. Validation of the experimental test as a predictor of on-job performance is a task that does not lie within the scope of this study, but which

Table 14

Ranges of Test Performances (IPR) and Navy General Classification Test (GCT) Scores

Pacing	Racial Group	GCT Group	GCT Low	GCT High	GCT Range	IPR Low	IPR High	IPR Range
SELF	Nonwhite	Low ($\leq$50)	32	49	17	97	503	406
		High ($\geq$50)	56	62	6	164	168	4
	White	Low ($\leq$50)	38	48	10	88	215	127
		High ($\geq$50)	51	70	19	123	388	265
MACHINE	Nonwhite	Low ($\leq$50)	39	49	10	123	240	117
		High ($\geq$50)	50	69	19	123	308	185
	White	Low ($\leq$50)	33	48	15	148	265	117
		High ($\geq$50)	50	75	25	117	321	204
COMBINED	Nonwhite	Low ($\leq$50)	32	49	17	97	503	406
		High ($\geq$50)	50	69	19	154	308	154
	White	Low ($\leq$50)	33	48	15	88	265	178
		High ($\geq$50)	50	75	25	117	388	271

Note. IPR presented in bits/sec x 10^3.

nonetheless is essential to determine the exact nature of the test's value in the selection process.

E. INTELLIGENCE, TEST PERFORMANCE, AND RACIAL DIFFERENCES

Analysis of overall performance during the test showed that, aside from the "learning effect" discussed above, only racial differences proved to be a significant factor in learning performance. This factor was further isolated to the machine-paced test groups only, where white performance exceeded nonwhite performance; the reverse was true for the self-paced group, although the difference in this case was not significant. It would seem that the pacing mode of the test itself affected the performance of the nonwhite subjects.

While no concrete theory is advanced here to explain this effect, it is possible that the pressure of machine-paced presentation made the nonwhite subjects more anxious, and thus less able to perform up to their true ability. This hypothesis is suggested by studies by Taylor and Spence (1952) and by Ramond (1953), in which subject anxiety was found to have a detrimental effect on performance in serial and verbal learning tasks.

In the light of in-house findings of racial test bias within the Navy BTB (Stephan, 1973; Thomas, 1972c), the results of a 2 X 2 Chi-square test of GCT score distribution by racial group within the test sample are of interest. When the sample was divided into four groups by race (white, nonwhite) and GCT score (below the mean score of 50, above 50), the cells showed 87 whites with GCT scores of 50 or above, 17

below 50, 9 nonwhites with scores 50 or higher, and 27 nonwhites below the GCT mean. This distribution yielded a Chi-square statistic (*df* = 1) of 35.33, which is significant at the .001 level of confidence. The men in the test sample were by no means evenly distributed across the GCT score range with respect to racial lines.

F. TESTING CONSIDERATIONS

Internal reliability of the test itself was established using the split-half technique and employing the Spearman-Brown coefficient. Reliability coefficients were, in general, overwhelmingly strong, especially when the total testing time (4 min.) is considered. In order to establish the validity of the test in predicting on-job training performance, comprehensive follow-up study is necessary. Most Navy classification tests in use today show strong predictive capability for school performance (Thomas, 1972a and 1972b), but are not validated against on-job performance.

While the apparatus assembled for this study was not practical for use in an operational setting, the test materials are readily adaptable to existing teaching machines, many of which could provide on-the-spot scoring as well. The test concept is also highly compatible with computer-based teaching systems, such as PLATO. Such a system would permit testing of potential recruits at remote terminal sites connected to a central computer system that would administer, score, and evaluate performances in a real-time setting.

VI. CONCLUSIONS

Conclusions that may be drawn from the results of this study are as follows:

1. Learning did, in fact, take place during the course of the administration of the experimental test.

2. Termination of testing prior to the "perfect learning" point is considered to give greater differentiation between individual performances, but the point chosen (after 10 trials) may or may not be optimal.

3. The rate at which information was processed by the self-paced group was not significantly different from that of the machine-paced groups, indicating that IPR is, in many cases, independent of presentation rate.

4. The self-paced task enables a superior performer to attain his own level without being limited by presentation rate.

5. No significant difference was seen between white and nonwhite performance in the self-paced task, but this was not true in the machine-paced situation, where white scores exceeded nonwhite.

6. Intra- and interpair similarity was not a factor in learning performance, suggesting that item similarity using these random shapes may not be important in constructing equivalent stimulus lists.

7. Performance on the experimental test did not correlate highly with performance on standard verbal-oriented intelligence tests (Navy GCT).

Appendix A

SUBJECTS' INSTRUCTIONS — EXPERIMENT I

Three columns of pairs of two-dimensional shapes are listed on the following pages. In each case, one shape has been arbitrarily selected as "correct." Selection of the "correct" shape in each pair was made without regard to any of the other selections in other pairs. That is, no systematic procedure was used.

(1) Go through the entire list, pair by pair, and circle the shape in each pair which you think is the one that was selected as "correct."

(2) Go through the entire list again, this time writing in the space between the members of each pair a number one (1), two (2), or three (3), as follows:

"1" - If the two shapes of the pair appear to you to be very similar.

"2" - If the two shapes of the pair appear to you to be only slightly similar.

"3" - If the two shapes of the pair appear to you to be dissimilar.

When you have completed these tasks, return this booklet to your instructor. Thank you for your time.

Appendix B

SUBJECTS' INSTRUCTIONS — EXPERIMENT II

I am asking you to volunteer to take an experimental test. If you do volunteer, the test itself will take about four minutes of your time. This test is completely experimental--nothing will ever get into your training jacket here at NTC, nor into your service record. You will be helping me to "test out the test."

In a few minutes, I will be showing you a lot of pairs of black shapes on a white background. These shapes were drawn by a computer, and they are not supposed to look like or mean anything in particular. The shapes will appear on this screen two by two. One of the shapes in each pair is one that I have decided to call the correct answer; the other shape is a wrong answer. When you first see a pair of the figures, I want you to guess which one is the one I called "correct." You will give your answers by pressing these two buttons. If you think the right-hand shape is "corect," press the right-hand button. If you think it's the left one, press the left button. If you guessed correctly, a light at the bottom of the screen will come on. If you made the wrong guess, no lights will light up. You will see each pair of shapes more than once. The first time a pair appears, you will be guessing at the correct shape, but I want you to try to remember which shape is correct, so you can make the right answer without guessing

the next time you see the pair again. There are a lot of different pairs, so this is a tough test. Don't worry if you don't seem to be getting a lot of right answers; just do your best and try to remember which shape is the correct one in each pair. Make only one answer each time; if you didn't get it right, you know it's the other one.

SPECIAL INSTRUCTIONS FOR SELF-PACED GROUP:

When you have made your answer, press one of these buttons on the sides of the box to go on to the next pair. Work as fast as you can, but don't rush it and don't just go through the whole time guessing.

SPECIAL INSTRUCTIONS FOR MACHINE-PACED GROUP:

The pairs will come on the screen every few seconds, so you don't have a lot of time to decide which one is correct. You <u>will</u> have plenty of time to get a good look at the shapes, make your decision, and make your answer. The pairs will be on the screen the same amount of time each time.

REMAINING INSTRUCTIONS FOR ENTIRE SAMPLE:

Remember that there are a lot of different pairs, but that you will see each pair several times. The correct shape in each pair will always be the correct answer, but will sometimes be on the right-hand side and sometimes on the left. It will always appear with the same shape. The shapes will not be turned around or flipped over--they will always be the same way as you first saw them.

Appendix C

GCT SCORES AND TEST PERFORMANCE BY TEST GROUP

TEST GROUP 1

Subject	GCT	IPR Block 1	IPR Block 2	IPR Block 3
1	58	388	365	410
2	59	261	341	369
3	48	105	176	199
* 4	40	76	152	152
5	62	216	288	306
6	57	170	170	170
* 7	41	181	237	237
8	59	239	279	259
9	55	175	219	219
*10	40	131	112	131
*11	42	125	187	200
*12	46	160	131	203
*13	62	156	173	173
*14	40	180	246	197
15	61	173	308	308
16	53	138	157	276
17	38	204	204	238
*18	32	248	248	223
*19	47	405	405	405
20	58	123	135	197
21	60	273	323	373
22	70	159	159	209
23	60	113	97	161
24	59	156	156	242
*25	49	426	503	581
26	56	208	208	340
27	51	130	162	260
28	72	71	107	85
29	63	169	192	203
*30	39	90	112	90
31	58	143	157	214
32	55	191	287	306
33	60	187	187	214
34	43	182	199	265
35	41	180	180	232

Note: (1) IPR is presented in bits/sec X 10^3.

(2) Nonwhite subjects designated by *.

TEST GROUP 2

Subject	GCT	IPR Block 1	IPR Block 2	IPR Block 3
* 1	39	148	240	240
2	56	240	333	333
* 3	44	166	240	296
* 4	52	166	166	148
5	51	185	277	277
6	65	185	240	222
7	65	203	296	333
* 8	40	222	185	259
9	57	203	314	314
10	45	166	259	259
11	53	240	314	296
*12	69	259	333	333
13	59	259	296	296
14	33	185	240	296
15	46	203	277	296
16	55	203	314	333
17	51	277	259	296
18	54	92	129	129
19	51	222	314	296
20	65	296	296	333
*21	46	129	203	259
*22	48	111	166	259
*23	53	55	148	259
24	37	185	203	259
25	35	129	92	222
*26	42	92	129	148
*27	45	240	166	203
28	61	222	314	333
*29	44	192	129	111
30	46	222	222	277
31	63	185	277	314
32	60	222	296	296
*33	40	166	222	333
34	40	148	92	222
35	56	277	333	314
36	65	129	240	277
37	66	185	314	277

Note: (1) IPR is presented in bits/sec X 10^3.

(2) Nonwhite subjects designated by *.

TEST GROUP 3

Subject	GCT	IPR Block 1	IPR Block 2	IPR Block 3
1	66	185	148	185
2	61	185	185	240
* 3	40	240	240	203
4	61	166	185	240
* 5	47	148	240	185
6	61	185	259	314
7	67	259	277	314
8	46	203	240	240
* 9	42	166	185	185
*10	59	240	259	277
11	65	203	259	277
12	68	185	240	314
13	58	148	259	277
14	59	166	259	240
*15	46	222	240	240
16	38	185	240	222
17	62	185	240	185
18	53	148	166	222
19	59	222	166	240
20	56	240	240	240
21	63	203	240	259
22	57	185	277	333
23	58	148	240	240
24	44	129	203	296
*25	59	240	203	296
26	62	185	185	240
27	75	166	222	166
28	60	259	314	333
*29	42	240	166	129
*30	49	148	129	185
31	59	166	314	314
32	54	185	259	296
33	63	185	333	314
34	55	240	314	314
35	63	166	185	296
36	58	240	296	314

Note: (1) IPR is presented in bits/sec X 10^3.

(2) Nonwhite subjects designated by *.

TEST GROUP 4

Subject	GCT	IPR Block 1	IPR Block 2	IPR Block 3
1	61	222	314	277
2	59	166	185	240
3	52	129	222	277
4	73	185	333	314
5	62	185	240	203
6	65	148	240	203
7	55	129	203	222
8	59	185	259	240
9	69	296	333	333
10	57	166	240	296
11	59	240	296	314
12	56	222	314	333
13	51	166	277	222
14	50	111	185	296
*15	--	296	259	203
16	61	259	259	277
17	64	166	222	277
18	54	129	240	222
19	63	203	222	185
20	56	148	259	296
21	47	277	240	277
*22	50	111	129	129
23	66	203	185	203
*24	52	259	259	314
25	72	203	203	277
26	48	185	277	277
27	52	203	203	222
28	60	222	314	314
29	59	129	166	203
30	53	185	240	296

Note: (1) IPR is presented in bits/sec X 10^3

(2) Nonwhite subjects designated by *.

REFERENCES

Arima, J. K. Verbal discrimination learning: An analysis of randomly presented 2-, 3-, and 4-word items. JSAS Catalogue of Selected Documents in Psychology, 1974, 4, 116.

Arima, J. K., & Gray, F. D. Information transfer in 2-, 3-, and 4-word verbal discrimination learning. Proceedings of the 20th International Congress of Psychology, 1972, 427. (a)

Arima, J. K., & Gray, F. D. Information analysis of 2-, 3-, and 4-word verbal discrimination learning. Proceedings of the 80th Annual Convention of the American Psychological Association, 1972, 869. (b)

Arnoult, M. D. Shape discrimination as a function of angular orientation of the stimuli. Journal of Experimental Psychology, 1954, 47, 323-328.

Arnoult, M. D. Familiarity and recognition of nonsense shapes. Journal of Experimental Psychology, 1956, 51(4), 269-276.

Attneave, F., & Arnoult, M. D. The quantitative study of shape and pattern perception. Psychological Bulletin, 1956, 53, 452-471.

Baltutis, J. S. Information transfer in 2-, 3-, and 4-word verbal discrimination learning with two stimulus presentation rates. M.S. Thesis, U.S. Naval Postgraduate School, Monterey, California, 1972.

Bugarin, T. E. Verbal discrimination learning in a random mixture of 2-, 3-, and 4-word items with two stimulus presentation rates. M.S. Thesis, U.S. Naval Postgraduate School, Monterey, California, 1973.

Cattell, R. B. Theory of fluid and crystallized intelligence: A critical study. Journal of Experimental Psychology, 1963, 54, 1-22.

Estes, W. K. Learning theory and intelligence. American Psychologist, 1974, 29, 740-749.

Fitts, P. M., Weinstein, M., Rappaport, M., Anderson, N., & Leonard, A. J. Stimulus correlates of visual pattern recognition: A probability approach. Journal of Experimental Psychology, 1956, 51, 1-11.

Fox, W. L., Taylor, J. E., & Caylor, J. S. Aptitude level and the acquisition of skills and knowledges in a variety of military training tasks (HumRRO Tech. Rep. 69-6). U.S. Army, HumRRO Div. #3, 1969.

Gibson, E. J. Intralist generalization as a factor in verbal learning. Journal of Experimental Psychology, 1942, 30, 185-200.

Gibson, E. J. A re-examination of generalization. Psychological Review, 1959, 66, 3040-3042.

Green, E. J., & O'Connell, J. A. An annotated bibliography of visual discrimination learning. Teachers College Press, Columbia University, New York, N. Y., 1969.

Gray, F. D. Information transfer in 2-, 3-, and 4-word verbal discrimination. M.S. Thesis, U.S. Naval Postgraduate School, Monterey, California, 1971.

Haber, R. N. How we remember what we see. Scientific American, 1970, 222(5), 104-112.

Hebb, D. O. Textbook of psychology. Philadelphia: W. B. Saunders, 1972.

Jensen, A. R. Social class, race, and genetics: Implications for education. American Educational Research Journal, 1968, 5, 1-42.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Wadsworth, 1968.

Matarazzo, J. D. Wechsler's measurement and appraisal of adult intelligence. Baltimore: Williams & Wilkins, 1972.

Ornstein, R. E. The psychology of consciousness. San Francisco: W. H. Freeman, 1972.

Ramond, C. K. Anxiety and task as determinants of verbal performance. Journal of Experimental Psychology, 1953, 46, 120-124.

Stephan, R. A. Evidence of racial bias in military testing. Paper presented at 31st meeting of Military Operations Research Society, Monterey, California, 1973.

Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. Journal of Experimental Psychology, 1952, 44, 61-64.

Thomas, P. J. The relationship between Navy classification test scores and final school grade in 104 Class "A" Schools (Res. Rep. SRR 72-15). San Diego: U.S. Navy Personnel & Training Research Laboratory, 1972. (a)

Thomas, P. J. The relationship between Navy classification test scores and final school grades in 98 Class "A" schools (Res. Rep. SRR 72-22). San Diego: U.S. Navy Personnel & Training Research Laboratory, 1972. (b)

Thomas, P. J. An investigation of possible test bias in the Navy Basic Test battery (Tech. Bull. STB 73-1). San Diego: U.S. Navy Personnel & Training Research Laboratory, 1972. (c)

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

INITIAL DISTRIBUTION LIST

		No. of Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
3.	Naval Postgraduate School Department of Operations Research and Administrative Sciences Monterey, California 93940	1
4.	Chief of Naval Personnel Pers 11b Department of the Navy Washington, D. C. 20370	1
5.	Assoc. Professor James K. Arima, Code 55Aa Department of Operations Research and Administrative Sciences Naval Postgraduate School Monterey, California 93940	1
6.	Asst. Professor Douglas Neil, Code 55Ni Department of Operations Research and Administrative Sciences Monterey, California 93940	1
7.	Man-Machine Systems Design Library Root 107 Code 55 Aa Naval Postgraduate School Monterey, California 93940	1
8.	LT Peter A. Young, USN Code 312 Navy Personnel Research and Development Center San Diego, California 92152	1